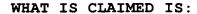
3



2 July 1/2 3

1\ A method comprising:

segmenting video data to create a video clip based on timing data; and

determining a most likely gesture in the video clip.

- 2. The method of claim 1, wherein determining includes
 determining a probability that each of a plurality of
 predefined gestures are performed in the video clip contains
 the predefined gesture.
 - 3. The method of claim 2, wherein determining the probability that the video clip contains each of the predefined gesture includes evaluations of Hidden Markov Models.
- 1 4. The method of claim 1, wherein the timing data includes beat data corresponding to a beat of audio data.
- The method of claim 4, further comprising:
 receiving the audio data; and
 extracting the beat data from the audio data.
- 1 6. The method of claim 4, wherein the video clip
 2 includes a portion of the video data corresponding to a
 3 predefined time window surrounding the occurrence of at least
 4 one beat.

3.

1

2

- The method of claim 1, further comprising displaying
 a target gesture to be performed by the subject of the video
- 3 data.
- 1 8. The method of claim 1, wherein each video clip 2 contains video frames.
- 9. The method of claim 1, further comprising identifying moving regions in each video frame in the video clip.
- 1 10. The method of claim 9, further comprising generating 2 a feature vector for each video frame of the video clip.
 - 11. The method of claim 1, further comprising generating a score based on whether the video clip contains the target gesture.
 - 12. The method of claim 11, further comprising displaying the score.
- 1 13. The method of claim 1, wherein determining if the
 2 video clip contains the predefined gesture includes generating
 3 a gesture probability vector having a plurality of elements,
 4 each element being associated with one of a plurality of
 5 predefined gestures and representing a probability that the

- 6 video clip contains each of the associated predefined
- 7 gestures.
- 1 14. A\system comprising:
- a temporal segmentor connected to receive video data and
- 3 to create a video clip from the video data based on timing
- 4 data; and
- 5 a recognition engine, in communication with the temporal
- 6 segmentor, to determine if the yideo clip contains a
- 7 predefined gesture.
- 1 15. The system of claim 14, wherein the recognition engine includes a plurality of Hidden Markov Models.
- 1 16. The system of claim 14, further comprising:
- a timing data source, in communication with the temporal
- segmentor, to provide the timing data to the temporal
- 4 segmentor; and
- a video source, in communication with the temporal
- 6 segmentor, to provide the video data $t \delta$ the temporal
- 7 segmentor.
- 1 17. The system of claim 14, further comprising a move
- 2 subsystem, in communication with the timing data source, to
- provide a target gesture to be performed by the subject of the
- 4 video data.

2

3

4

1

3

1

2

3

4

5

6

- 1 18 The system of claim 17, wherein the target gesture 2 is a dance move that is to be performed by the subject of the 3 video data.
 - 19. The system of claim 17, further comprising a scoring subsystem, in communication with the recognition engine and the move subsystem, to determine if the video clip contains the target gesture.
- 20. The system of claim 19, further comprising a display subsystem, in communication with the scoring subsystem, to display a score that is a function of whether the video clip contains the target gesture.
 - 21. The system of claim 20, wherein the display subsystem is in communication with the move subsystem and is configured to display a gesture request based on the target gesture.
 - 22. The system of claim 14, wherein the recognition engine is configured to recognize predefined gestures and to produce a gesture probability vector having elements, each element being associated with one of the predefined gestures and representing the probability that the video clip contains the associated predefined gesture.

- 1 23. The system of claim 14, wherein the timing data 2 source includes:
- an audio source that provides an audio data; and
- a beat extractor, in communication with the audio source,
- 5 that extracts beat data from the audio data.
- 1 24. The system of claim 23, wherein the video clip 2 corresponds to a beat in the beat data.
- 25. The system of claim 24, wherein the video clip includes a portion of the video data corresponding to a predefined time window surfounding the occurrence of the beat.
- 26. A computer program product, tangibly stored on a computer-readable medium, for recognizing gestures contained in video data, comprising instructions operable to cause a programmable processor to:
- segment the video data to create a video clip based on timing data; and
- 7 determine if the video clip contains a predefined 8 gesture.
- 27. The product of claim 26, further comprising
 instructions operable to cause the programmable processor to:

extract beat data from an audio signal; and

- segment the video data to create the video clip using the beat data. 5 An audio-visual processing system including: 1 a video source to provide video data; 2 an audio source to provide audio data; 3 a speaker to play at least a portion of the audio data; 5 and a computer program product, tangibly stored on a 6 computer-readable medium, for recognizing gestures contained 7 in video data, comprising instructions operable to cause a 8 programmable processor, in communication with the video source 9 and the audio source, to: 10 extract beat data from the audio data; 11 segment the video data to create a video clip based 12 on beat data; and 13 determine if the video clip contains a predefined 14 gesture. 15 The video processing system of claim 28, wherein the 1 29. computer program product further includes instructions 2
- perform a Hidden Markov Model process to determine if the video clip contains the predefined gesture.

operable to cause the programmable processor to:

- 1 30. The video processing system of claim 28, further
- 2 comprising a display to display information based on whether
- 3 the video clip contains the predefined gesture.